Published in final edited form as:

Neuroimage. 2013 May 1; 71: 224–232. doi:10.1016/j.neuroimage.2013.01.014.

A Two-level Multimodality Imaging Bayesian Network Approach For Classification of Partial Epilepsy: Preliminary Data

Susanne G. Mueller¹, Karl Young¹, Miriam Hartig¹, Jerome Barakos², Paul Garcia³, and Kenneth D. Laxer²

¹Center for Imaging of Neurodegenerative Diseases and Dept. of Radiology, University of California, San Francisco

²Sutter Pacific Epilepsy Program, California Pacific Medical Center, San Francisco

³Department of Neurology, University of California, San Francisco

Abstract

Background—Quantitative neuroimaging analyses have demonstrated gray and white matter abnormalities in group comparisons of different types of non-lesional partial epilepsy. It is unknown to what degree these type-specific patterns exist in individual patients and if they could be exploited for diagnostic purposes. In this study, a two-level multi-modality imaging Bayesian network approach is proposed that uses information about individual gray matter volume loss and white matter integrity to classify non-lesional temporal lobe epilepsy with (TLE-MTS) and without (TLE-no) mesial-temporal sclerosis and frontal lobe epilepsy (FLE).

Methods—25 controls, 19 TLE-MTS, 22 TLE-no and 14 FLE were studied on a 4T MRI and T1 weighted structural and DTI images acquired. Spatially normalized gray matter (GM) and fractional anisotropy (FA) abnormality maps (binary maps with voxels 1 SD below control mean) were calculated for each subject. At the first level, each group's abnormality maps were compared with those from all the other groups using Graphical-Model-based Morphometric Analysis (GAMMA). GAMMA uses a Bayesian network and a Markov random field based contextual clustering method to produce maps of voxels that provide the maximal distinction between two groups and calculates a probability distribution and a group assignment based on this information. The information was then combined in a second level Bayesian network and the probability of each subject to belong to one of the three epilepsy types calculated.

Results—The specificities of the two level Bayesian network to distinguish between the three patient groups were 0.87 for TLE-MTS and TLE-no and 0.86 for FLE, the corresponding sensitivities were 0.84 for TLE-MTS, 0.72 for TLE-no and 0.64 for FLE

Conclusion—The two-level multi-modality Bayesian network approach was able to distinguish between the three epilepsy types with a reasonably high accuracy even though the majority of the images were completely normal on visual inspection.

© 2012 Elsevier Inc. All rights reserved.

Corresponding Author: Susanne G. Mueller MD, Center for Imaging of Neurodegenerative Diseases, Department of Veterans Affairs (DVA) Medical Center, Clement Street 4150, San Francisco, CA, 94121, Phone: 415 221 4810 ext 2538, Fax: 415 668 2864, susanne.mueller@ucsf.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

TLE; FLE; classifier; Bayesian network; epilepsy; DTI; gray matter volume

INTRODUCTION

Quantitative unbiased whole brain neuroimaging analyses have demonstrated subtle but widespread gray and white matter abnormalities in different types of non-lesional partial epilepsy even in the absence of macroscopic lesions. It is generally assumed that the structural abnormalities in the focus are caused by microscopic malformations and/or excitotoxic neuronal damage while abnormalities beyond the focus are the result of excitotoxic damage due to spreading epileptogenic discharges and/or deafferentation (Bernhardt et al. 2008; Mueller et al., 2009 a; Concha et al., 2009; Bonilha et al. 2006). The distribution of these abnormalities is determined by the location of the focus and the brain regions connected to it and therefore will be different for different epilepsy types. For example, temporal lobe epilepsy with mesial temporal sclerosis (TLE-MTS) is characterized by bilateral, ipsilaterally pronounced hippocampal and mesial temporal gray matter volume loss that extends into posterior temporo-parietal and frontal lobe regions while TLE with normal MRI (TLE-no) shows less well lateralized temporo-lateral volume losses extending into the insula and frontal lobes (Mueller et al., 2009a). Both are associated with temporal and frontal white matter abnormalities (Concha et al., 2009). The seizure semiology of frontal lobe epilepsy (FLE) is clinically more variable than that of TLE. Furthermore, frontal lobe regions are densely interconnected between themselves and with extrafrontal regions that allows a variety of pathways for seizure spread. Therefore, gray matter abnormalities in FLE are more heterogeneous and in group comparisons less pronounced in the frontal lobe but more prominent in subcortical relay structures, e.g. thalamus (Salmenpera et al., 2007, Widjaja et al., 2011).

Given the characteristic differences at the group level, the question arises to what degree these type-specific patterns exist in individual patients and if they could be exploited for diagnostic purposes. A precondition to answer these two questions though, is the availability of a method that reliably identifies these abnormalities in individual subjects. Currently, the most common strategies to characterize brain structural abnormalities in individual subjects are either to use an imbalanced (single patient vs. control group) version of the general linear model or to employ a thresholding approach, e.g., by generating z-score maps. Both approaches are highly dependent on subjectively chosen thresholds and the former method also violates the basic assumptions of parametric statistical mapping. The diagnostic decision finally is typically made by comparing the extent and distribution of the abnormalities found in the group comparison with those in the single subject analysis either by visual inspection or by calculating a measure of overlap. Unfortunately, the between-subject variation of brain structural abnormalities even within well defined and seemingly uniform patient populations complicates interpretation of these comparisons. If information from two or more imaging modalities has to be integrated, this process gets even more complex.

The objective of this study was therefore to develop a robust and unbiased multimodality (volumetric MR and DTI), whole brain, single subject analysis approach that assesses to what degree the characteristic brain gray and white matter abnormalities found in group comparisons in non-lesional TLE with and without MTS and FLE are present in individual patients and if that information can be used to distinguish between the three epilepsy types. To that purpose a two-level Bayesian network approach was chosen. At the first level, Graphical-Model-based Morphometric Analysis (GAMMA (Cheng and Herskovits, 2012))

was used to determine the characteristic brain structural abnormalities for each of the three epilepsy types. GAMMA uses a Bayesian network approach to detect linear and non-linear associations between voxel clusters and a function variable, e.g. group, in each of the modalities. The second level Bayesian network used the combined information from the first level to calculate the probability of an individual patient belonging to each of the three epilepsy types. It was expected that a network using two modalities would have a higher classification accuracy than a single modality network.

METHODS

Study Population

The committees of human research at the University of California, San Francisco (UCSF), California Pacific Medical Center, San Francisco (CPMC) and VA Medical Center, San Francisco approved the study, and written informed consent was obtained from each subject according to the Declaration of Helsinki. The study population consisted of 80 subjects: 25 controls (34.2 ± 15.6 , female/male: 18/7), 19 patients suffering from TLE with unilateral mesial temporal seizure origin and ipsilateral mesial-temporal sclerosis (TLE-MTS) (mean age: 40.8 ± 10.9 , female/male: 11/8, left/right focus: 12/7, mean age at onset: 10.8 ± 9.1 , mean epilepsy duration: 31.2 ± 12.7 years), 22 patients suffering from TLE with unilateral mesial-temporal seizure origin and normal MRI (TLE-no) (39.2 ± 10.4 , female/male: 16/6, left/right focus: 14/8, mean age at onset: 24.6 ± 11.7 , mean epilepsy duration: 15.4 ± 12.7 years) and 14 patients suffering from non-lesional frontal lobe epilepsy (FLE) (mean age: 27.3 ± 9.8 ; female/male: 7/7, left/right/bilateral: 6/4/4, mean age at onset: 15.4 ± 8.4 , mean epilepsy duration: 13.0 ± 8.7 years). The identification of the epileptogenic focus was based on seizure semiology and prolonged ictal and interictal Video/EEG/Telemetry (VET) in all patients; the presence/absence of MTS in TLE was based on hippocampal subfield volumetry (Mueller et al., 2009 b). All patients reported having been seizure free for at least 24 h before the 4T study. FLE were significantly younger than TLE-MTS and TLE-no. TLE-no were significantly older at the onset of their epilepsy than TLE-MTS who had a longer duration of epilepsy than TLE-no and FLE.

MR Imaging

All studies were performed on a Bruker MedSpec 4T system controlled by a Siemens Trio™ console and equipped with a USA instruments eight channel array coil. The following sequences, which were part of a larger research imaging and spectroscopy protocol, were acquired: 1) T1-weighted whole brain gradient echo MRI TR/TE/TI = 2300/3/950 ms, $1.0 \times 1.0 \times 1.0 \text{ mm}^3$ resolution (for tissue segmentation) 2) 3D T2-weighted turbo spin-echo sequence, TR/TE = 3500/356 ms, $1.0 \times 1.0 \times 1.0 \text{ mm}^3$ resolution (for calculation of intracranial volume). 3. high resolution T2 weighted fast spin echo sequence for hippocampal subfield volumetry (TR/TE: 3500/19 ms, $0.4 \times 0.4 \text{ mm}$ in plane resolution, 2 mm slice thickness, 24 interleaved slices, angulated perpendicular to the long axis of the hippocampal formation. 4. EPI-based DTI (TR/TE = 6000/77, $2 \times 2 \times 3 \text{ mm}^3$ resolution, 6 diffusion encoding directions with $b = 800 \text{ s/mm}^2$, repeated 4 times to boost SNR. Total acquisition time for structural MRIs: ~ 30 min.

Post-processing and Image Analysis

Post-processing—The T1 weighted image was segmented into tissue categories (gray matter (GM), white matter and CSF) using the Expectation Maximization Segmentation algorithm (van Leemput et al., 2003) as implemented in SPM2. The DTI data was motion and eddy current corrected using the FLIRT and FUGUE algorithms from FSL (<http://www.fmrib.ox.ac.uk/fsl>) and an additional geometric distortion correction performed (Tao et al., 2009). The DTI images were then co-registered to the T1 image and up-sampled to the

T1 resolution. Fractional anisotropy (FA) maps were calculated using the *teem* algorithms (teem.sourceforge.net). Unbiased project specific, symmetrical atlases (GM, FA) were generated from the original and side flipped control images using the fast, high degree diffeomorphic image registration algorithm (DARTEL) implemented in SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>). Images from patients with right sided-focus were side-flipped to ensure that the seizure onset was on the same side in all patients with lateralized seizure onset. The original and side-flipped images from all subjects were warped onto the corresponding atlas using DARTEL. The spatially normalized GM maps were multiplied with the Jacobian determinants of the warping step to preserve the original volume and the intensity of the resulting maps were corrected for differences in intracranial volume (ICV). The spatially normalized subject images (GM, FA) were smoothed with an 8 mm FWHM kernel.

First level Bayesian Network: Graphical-Model-based Morphometric Analysis

—The smoothed spatially normalized original and side flipped control images were used to calculate a voxel-wise mean and standard deviation (SD) map for each image modality. Each subject's smoothed GM map and FA map was compared against the corresponding mean map and binary maps indicating voxels with intensities at least one SD below the mean generated. The threshold was intentionally chosen well within the control range to enhance the possibility to identify regions showing mild but consistent abnormalities in the patient groups. The binary maps were used as input for Graphical-Model-based Morphometric Analysis (GAMMA) (Chen and Herskovits 2005; Chen and Herskovits 2007, Chen and Herskovits, 2012). GAMMA uses a Bayesian network to represent associations among voxels and a function variable, e.g., group, and applies a Markov random field based contextual clustering method to identify clusters of voxels based on their spatial relationship and similar association with the function variable. The stability of the association between these voxels and the function variable is determined using ensemble learning with bootstrap aggregation. The result is a label map (binary image of the voxel subset associated with the function variable) and a belief map (label map weighted by the confidence in the voxel/function variable association). GAMMA also determines each subject's group membership using a regional state inference algorithm. GAMMA maps were calculated for the following comparisons: TLE-MTS less than controls (GM, FA), TLE-no less than controls (GM, FA), FLE less than controls (GM, FA), TLE-MTS less than TLE-no (GM, FA), TLE-MTS less than FLE (GM) and more than FLE (FA), TLE-no less than FLE (GM) and more than FLE (FA) and MTS less than non-MTS (GM). All labels were based on whole brain binary maps except for the MTS vs non-MTS label which was generated by comparing binary maps restricted to the bilateral hippocampus – entorhinal cortex – parahippocampus region of TLE-MTS with those of non-TLE-MTS (controls, regular TLE-no, regular FLE). The direction of the comparisons, e.g. TLE-MTS has less GM than FLE, was confirmed by testing the two possible contrasts in SPM8 ($p < 0.005$ uncorrected).

The correspondence between the label maps and individual binarized maps of subjects not included in the initial label generation (e.g. correspondence of the binarized map of a FLE patient with the TLE-MTS > TLE-no label maps) was determined by extracting voxels represented in both maps, multiplying them with the intensities of the corresponding voxels in the belief map and calculating the sum of the their intensities (weighted correspondence). The subject classification provided by GAMMA for the initial comparison and the weighted correspondence was used to determine the label map specific classification threshold. The individual weighted correspondence and the label specific threshold was then used to classify subjects not included in the initial group comparison (at or above label threshold: positive; below label threshold: negative).

Second Level Bayesian Network—The purpose of this step was to combine the information obtained from the two modalities and the different comparisons at the first level. Expert knowledge about the different groups and their imaging characteristics was used to model the causal relationships and build a network structure using the Graphical Network Interface software (GeNIe, <http://genie.sis.pitt.edu>) (cf. Figure 1). In addition to the nodes using imaging information, one node using clinical information, i.e., the probability of having seizures or not, was also included into the network, to distinguish epilepsy subjects with mild imaging abnormalities from controls with mild imaging abnormalities. Evidence nodes (yellow in Figures 1, 2a b) were modeled as uncertainty nodes (using the probabilities calculated by GAMMA)) and all intermediate (orange nodes in Figure 1, Figure 2a, b, Supplementary Figure and Table 1) and query nodes (dark orange nodes in Figure 1, Figure 2a, b) were modeled as deterministic nodes. The clustering algorithm as implemented in GeNIe was used for belief updating. The network allowed to directly address the following questions of interest: 1. The probability of a subject to be a control (no seizures, no structural abnormalities), to be partial epilepsy (seizures and structural abnormalities), or to be partial epilepsy with normal MRI (seizures, structural abnormalities not exceeding those in controls). 2. The probability of a subject classified as having partial epilepsy to be either TLE or FLE based on the pattern of the structural abnormalities. 3. The probability of a subject classified as TLE to be TLE-MTS or TLE-no based the pattern of the structural abnormalities. A more detailed description of the process can be found in the text to Supplementary Figure/Text 1. At each level, the option with the highest probability was chosen to represent the classification of the individual subject by the Bayesian network. Figure 1 displays the final network. In order to test the assumption that the multi-modality network achieved a higher classification accuracy than each of the modalities alone, two smaller networks were developed that used information from either GM or FA maps but otherwise had the same general network structure and prior knowledge as the two-modality version (cf. Figures 2a, b)

Statistical Analysis

In order to investigate to what degree these maps capture different abnormality patterns, the similarity between them was assessed by calculating the dice coefficient from the belief maps with a. all voxels (overlap), and b. thresholded so that only 15% of the high belief voxels were included (high belief overlap). The former threshold was chosen to obtain a measure of the total overlap of the different maps and the latter to obtain a measure of the overlap of the regions with the highest discriminability. The sensitivity and specificity of the two level Bayesian network to distinguish between the three different epilepsy types were calculated by comparing it with the clinical diagnosis based on VET and imaging. Only subjects that had imaging abnormalities exceeding those found in controls in at least one imaging modality were included in the sensitivity/specificity calculations. In the remainder, the imaging findings were considered to be non-diagnostic.

RESULTS

Group Differences as detected by GAMMA

Figure 3a depicts the gray matter belief maps generated by each comparison. Compared to controls, TLE-MTS had ipsilaterally pronounced bilateral hippocampal and ipsilateral insular and thalamic volume losses that were associated with more symmetrical bilateral dorso-lateral frontal and cerebellar gray matter losses. TLE-no had bilateral superior temporo-lateral, insular, opercular, thalamic and dorso-lateral frontal gray matter losses while FLE had bilateral insular and thalamic volume losses and comparatively mild, diffuse frontal lobe gray matter losses. As to be expected the differences between the three epilepsy types were less prominent. The most prominent differences between TLE-MTS and TLE-no

were the ipsilateral hippocampal and bilateral cerebellar volume losses and a more pronounced dorsolateral frontal volume loss in the TLE-MTS group. The differences between TLE-MTS and FLE were similar as those between TLE-MTS and TLE-no except that TLE-MTS had additional ipsilateral temporo-lateral volume losses and less pronounced dorso-lateral frontal volume losses. Finally, compared to FLE, TLE-no had more pronounced bilateral temporo-lateral and insular volume losses. The dice coefficients for the gray matter belief maps at the two thresholds ranged between 0.01 – 0.47 for the overlap maps and between 0.00 – 0.39 for the high belief overlap maps. The TLE-MTS vs control belief map had the largest mean dice co-efficient, i.e., the distribution of the abnormalities in this map had the highest concordance with the abnormalities in other maps. Please see also Table 1a. for all comparisons.

Figure 3b depicts the FA belief maps for each of the comparisons. In comparison to controls, TLE-MTS had lower FA in the corpus callosum, fornix, ipsilateral temporal stem and internal capsule. In accordance with the gray matter abnormality, FA decreases in TLE-no were less lateralized and most prominent in the internal capsule, centrum semiovale and genu of the corpus callosum. FLE had the most extensive FA abnormalities of the three epilepsy types with the most prominent decreases in the body of the corpus callosum, white matter of the parahippocampus, temporal stem, centrum semiovale, capsula interna/externa. As expected, TLE-MTS had more pronounced damage in ipsilateral parahippocampal, temporal stem and capsula internal white matter regions than TLE-no. The FA decreases in the comparisons FLE vs. TLE-MTS and FLE vs TLE-no showed a similar distribution as seen in the comparison FLE vs controls. The dice coefficients for the FA belief maps at the two thresholds ranged between 0.02 – 0.46 for the overlap maps and between 0.00 – 0.26 for high belief overlap maps. The FLE vs. TLE-no belief map had the largest mean dice co-efficient, i.e., the distribution of the abnormalities in this map had the highest concordance with the abnormalities in other maps. Please see also Table 1b for all comparisons.

Classification by two-level Bayesian network

Table 2 provides a comparison of the classification accuracy of each modality alone and the multi-modality network. In general the percentage of patients that could not be classified because the imaging abnormalities did not exceed that found in controls was higher for the two single modality networks, as was the percentage of incorrectly classified patients. The percentage of correctly classified patients was higher in the multi-modality version. The percentage of TLE-MTS and TLE-no correctly classified by GM abnormalities alone was higher than the percentage correctly classified by FA abnormalities alone. In contrast, the percentage of FLE that was correctly classified by FA abnormalities alone was slightly higher than that classified by GM abnormalities alone.

The two-level multi-modality Bayesian network that used the information from both modalities classified of 48 the 55 patients, 7 could not be classified because they had structural abnormalities not exceeding that observed in controls. Seventy-five percent of those who could be classified by the Bayesian network were correctly identified and 25% were incorrectly classified. Sixteen (84.2%) TLE-MTS were correctly identified as TLE-MTS. The remaining TLE-MTS were classified as FLE. The mean and standard deviation of the probability to be TLE-MTS of those correctly identified was 0.87 ± 0.14 . The mean probability of those TLE-MTS who were incorrectly classified by the Bayesian network as another epilepsy type was 0.90 ± 0.03 . Four (18.2%) TLE-no had no structural abnormalities exceeding that observed in controls and thus could not be classified by the Bayesian network approach. Of the remaining TLE-no, 13 TLE-no (72.2%) were correctly identified by the Bayesian approach; three were identified as TLE-MTS based on a positive MTS label and two as FLE. The mean and standard deviation of the probability to be TLE-no of those correctly identified was 0.76 ± 0.08 . The mean probability not to be TLE-no of those TLE-no

not recognized by the Bayesian network was 0.79 ± 0.11 . Three FLE had structural abnormalities within the normal range and were not classified by the two-level Bayesian network approach. Of the remaining FLE, seven were classified as FLE (63.6%) and four as TLE-no (36.4%). The mean and standard deviation of the probability to be FLE of those correctly identified was 0.87 ± 0.04 . The mean probability not to be FLE of those FLE not recognized by the Bayesian network was 0.72 ± 0.06 . Based on these observations, the sensitivity of the two level Bayesian network for partial epilepsy with structural abnormalities is 0.84 for TLE-MTS, 0.72 for TLE-no and 0.64 for FLE, the corresponding specificities are 0.87 for TLE-MTS and TLE-no and 0.86 for FLE.

Finally, as mentioned in “Methods”, the information about the probability of a subject having seizures or not was included in the second level Bayesian network since a distinction between controls and epilepsy patients solely based on imaging criteria was not an aim of the study. Including this information resulted in a 100% correct distinction between controls and patients. If the seizure evidence node would have been set to “seizure yes” for controls, 32% of the controls who had at least one positive label would have been identified as partial epilepsy (7 TLE-no, 1 TLE-MTS), the remainder would have been identified as partial epilepsy with imaging abnormalities not exceeding those in controls (all labels negative).

DISCUSSION

This study had the following main findings. 1. The two-level multi-modality Bayesian network using the information about individual patterns of gray and white matter structural abnormalities correctly identified the type of partial epilepsy, i.e., TLE-MTS, TLE-no or FLE, in 75% of the patients with diagnostic (GM and FA abnormalities exceeding those in age matched controls) MR exams. 25% of the patients were not correctly identified by that approach. As expected, the classification accuracy of the multi-modality network was higher than that of the single modality networks. 2. The classification accuracy depended on the epilepsy type and was highest in TLE-MTS and lowest in FLE. This reflects the heterogeneity of the three epilepsy types since the FLE group consisted of patients with non-lesional MRI but different localization of the epileptic focus within the frontal lobe. Taken together, the findings of this study suggest that the distribution patterns of gray and white matter structural abnormalities in the three most common types of partial epilepsy are sufficiently different that they can provide important diagnostic information even in patients whose MRI is normal on visual inspection.

Of the three different epilepsy types investigated in this study, TLE-MTS was not only the epilepsy type in which the classification accuracy of the two-level multimodal Bayesian network was the highest (84.2% correct) but it was also the only type in which a classification of all patients was possible because all had GM and FA abnormalities exceeding the amount seen in the control group. This is consistent with the fact that in contrast to the other two epilepsy types, the ipsilateral hippocampal atrophy that is the hallmark of TLE-MTS and an accompanying ipsilateral temporal lobe atrophy are typically already detectable on visual inspection and that these visible alterations are often accompanied by widespread gray and white matter damage beyond the temporal lobes in quantitative analyses. Accordingly both (GM, FA) TLE-MTS vs Control comparisons had a high number of discriminating voxels. A comparison of the classification accuracy of the single modality two level Bayesian networks showed that gray matter abnormalities allowed for a better discrimination of TLE-MTS from the other two types despite the higher dice coefficients with the GM belief maps than the FA belief maps (cf. Table 1a, b) indicating a higher degree of similarity with the GM abnormalities found in TLE-no and FLE. A potential explanation for the better classification performance of the TLE-MTS GM belief

maps could be that they all included the hippocampus, i.e. the region with the most prominent abnormality in this epilepsy type.

The two-level multi-modality Bayesian network approach correctly identified 72.2% of the TLE-no patients whose abnormalities exceeded those found in controls. A comparison of the performance of the two single-modality networks showed that similarly as in TLE-MTS, GM abnormalities identify TLE-no with a higher accuracy than FA abnormalities. A miss-classification of the TLE-no as TLE-MTS was equally common as the miss-classification as FLE. This finding and the similar dice coefficients for the overlap TLE-no vs. controls belief map with the TLE-MTS vs controls and the FLE vs controls belief maps indicates that the amount of similarity shared between TLE-no and TLE-MTS does not exceed that shared between TLE-no and FLE and thus supports the notion that TLE-no represents a different form of epilepsy and not just a milder version of TLE-MTS. About 18% of the TLE-no however had GM and FA abnormalities that were within the control range and thus without diagnostic information. This indicates that the TLE-no group consists of at least two subgroups: one that is associated with minimal GM and FA abnormalities and another subgroup that is associated with widespread GM abnormalities but relatively restricted FA abnormalities (cf. Table 1a, b. Figures 3a, b). The group with minimal structural abnormalities did not differ from the group with widespread GM abnormalities regarding age at onset (minimal: mean age at onset: 15.75 ± 6.65 , widespread: 26.56 ± 11.8 years, Mann-Whitney $p = 0.13$) or duration of epilepsy (minimal: mean duration: 18.61 ± 5.55 , widespread: $14.69.56 \pm 13.77$ years, Mann-Whitney $p = 0.39$) indicating that these two different TLE-no subgroups cannot just be explained by differences of epilepsy severity.

The lowest classification accuracy had the two-level Bayesian network approach in FLE in whom only 63% of the subjects with abnormalities exceeding those in controls were correctly classified. The most important factor contributing to this relatively low accuracy is that only 14 subjects of our current 4T cohort fulfilled the inclusion criteria of having a normal MRI and a well defined frontal lobe focus based on the results of the presurgical evaluation. This sample size is probably too small to truly capture the variability of such a complex syndrome as FLE, particularly since it also included three subjects in whom the amount of structural abnormalities did not exceed that of controls and whose inclusion in the label generation might even have resulted in an underestimation of the structural abnormalities. On the other hand, the fact that even in our small FLE population 21% of the subjects had no detectable structural abnormalities could indicate that there exists a sizable proportion of FLE patients that has either structural abnormalities of a type that cannot be detected by the techniques used in this study or whose brains are indeed structurally normal. As in the TLE-no group, differences of epilepsy duration alone are probably not sufficient to explain the absence of structural abnormalities in these three patients (minimal: mean age at onset: 15.33 ± 8.08 , widespread: 15.36 ± 8.87 years, minimal: mean duration of epilepsy: $12.56.56 \pm 5.25$, widespread: 13.09 ± 9.69 years)

Altogether 25% of the patients were miss-classified by the two-level Bayesian network approach. To some degree this can be explained by the fact that the number of patients was rather small and probably not sufficient to capture the variations of structural abnormalities in these three epilepsy types in its entirety. However, an alternative explanation is that there exist patients with discordant imaging and EEG findings and that this has a biological implication. For example, it is conceivable that a TLE-MTS patient with strong imaging features of FLE has a different outcome after anterior temporal lobe resection than a TLE-MTS patient with the typical TLE imaging features.

Finally, it should be mentioned that without the information regarding presence/absence of seizures a third of the controls group would have been identified as TLE-no by the two-level

Bayesian network approach. This however was intentionally so implemented into the second level a priori knowledge to account for the often rather subtle abnormalities in TLE-no (Mueller et al., 2007; Mueller et al., 2008; Bernhardt et al., 2010). In a larger patient population it might have been possible to further subdivide each group based on the severity of the brain structural abnormalities into mildly and severely affected subjects and by defining each of the subgroups separately in GAMMA to obtain a better distinction between controls and patients with mild abnormalities solely based on imaging information.

This study has several limitations. 1. The most important limitation is that it was not possible to validate the performance of the two-level Bayesian network in an independent patient population with non-lesional TLE-MTS, TLE-no and FLE since all patients of the 4T cohort fulfilling the inclusion criteria had been used to define the first level of the network. Therefore, the performance was tested on a group of 16 TLE and FLE patients who did not meet the inclusion criteria, either because of additional macroscopic lesions, e.g., frontal or temporal cortical malformations (TLE-L, FLE-L), bilateral or contralateral MTS (TLE-MTS plus), or widespread seizure onsets, e.g. lateral temporal or temporal-parietal instead of mesial temporal. It was expected that the two level Bayesian network would perform near chance level if the first level labels indeed capture features that are specific for non lesional mesial TLE and FLE. Seven or 43.8% of these atypical/lesional patients were correctly identified if the correct distinction between FLE and TLE and the correct identification of presence/absence of MTS were used as criteria to define a correct description by the network. Five of the correctly identified patients were TLE-MTS plus, i.e., could be expected to have very similar characteristics as TLE-MTS. 2. The method described in this paper is only one of several classification techniques, e.g. Focke et al (2011) and Keihaninejad et al (2012) used support vector machines to identify imaging features that distinguish between controls and TLE with and without MTS. The focus of the current paper is on the description of the method and thus we did not compare it to some of the other approaches but this is certainly something that needs to be done in the future if the Bayesian network approach is ever to be used in a clinical setting. 3. Prior knowledge and deterministic nodes were used to build the second level Bayesian network because of the small study population. This might explain why the probabilities of the correctly and incorrectly classified patients were within the same range. In a larger study population it is possible to use an Expectation Maximization (EM) algorithm to calculate the probability distributions of the intermediate and query nodes which should not only result in more realistic probability distributions at the level of the query nodes but might also better account for “unexpected” label combinations. Based on the experiences with this study, we estimate that the ideal training population for the development of two-level multi-modality classifier that adequately describes the variability within each epilepsy type and also allows the use of the EM algorithm would at least encompass about 40 TLE-MTS, 40 TLE-no and 50 FLE. Although age affects GM volume and FA, it was not attempted to account for GM/FA abnormalities due to age effects in the network because there were no significant age differences between controls and the three patient groups. However, FLE patients were younger than TLE and it cannot be excluded that this resulted in an underestimation of GM abnormalities in FLE in the TLE vs FLE comparisons and contributed to the lower classification accuracy in this group. 4. The small patient population did not allow to build separate networks for each hemisphere and thus the focus lateralization had to be known.

In conclusion, the multi-modality two-level Bayesian network approach described in the paper was able to distinguish between three types of non-lesional partial epilepsy with a reasonably high accuracy that likely can be further increased e.g., by studying larger patient populations and/or by adding information from other imaging modalities, e.g. PET, or EEG. In addition to showing promise as a diagnostic tool, it also provided important information about the nature of partial epilepsy by demonstrating that even a highly selected and

clinically homogeneous appearing patient group such as TLE-no is associated with a range brain structural abnormalities from very mild to severe. Based on that further refinement of the current network to account for subtypes within an epilepsy type and a confirmation of its performance in an independent population of TLE and FLE will be necessary before this approach can be used in the clinic.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the NIH grant RO1-NS31966 to K.D.L. We would also like to express our gratitude to Dr. R. Chen for the University of Pennsylvania who kindly provided expert advice in the use of GAMMA.

References

- Bernhardt BC, Worsely KJ, Besson P, Concha L, Lerch JP, Evans AC, Bernasconi N. Mapping limbic network organization in temporal lobe epilepsy using morphometric correlations: Insights into the relation between mesiotemporal connectivity and cortical atrophy. *Neuroimage*. 2008; 15:515–524. [PubMed: 18554926]
- Bernhardt BC, Bernasconi N, Concha L, Bernasconi A. Cortical thickness analysis in temporal lobe epilepsy. *Neurology*. 2010; 74:1776–1784. [PubMed: 20513813]
- Bonilha L, Rorden C, Appenzeller S, Coan AC, Cendes F, Li ML. Gray matter atrophy associated with duration of temporal lobe epilepsy. *NeuroImage*. 2006; 32:1070–9. [PubMed: 16872843]
- Concha L, Beaulieu C, Collins DL, Gross DW. White matter diffusion abnormalities in temporal lobe epilepsy with and without mesial temporal sclerosis. *J Neurol Neurosurg Psychiatry*. 2009; 80:312–319. [PubMed: 18977826]
- Chen R, Herskovits EH. Graphical-Model-Based Morphometric Analysis. *IEEE Trans Med Imaging*. 2005; 24:1237–1248. [PubMed: 16229411]
- Chen R, Herskovits EH. Graphical-model-based multivariate analysis of functional magnetic resonance data. *Neuroimage*. 2007; 35:635–647. [PubMed: 17258473]
- Chen R, Herskovits EH. Graphical model based multivariate analysis (GAMMA), an open source, cross-platform neuroimaging data analysis software package. *Neuroinformatics*. 2012; 10:119–127. [PubMed: 21882083]
- Focke NK, Yogarajah M, Symm MR, Gruver O, Paulus W, Duncan JS. Automated image classification in temporal lobe epilepsy. *Neuroimage*. 2012; 59:356–362. [PubMed: 21835245]
- Keihaninejad S, Heckemann RA, Gousias IS, Hajnal JV, Duncan JS. Classification and lateratization of temporal lobe epilepsies with and without hippocampal atrophy based on whole-brain automatic MRI segmentation. *PLoS ONE*. 7(4):e33096.10.1371/journal.pone.0033096 [PubMed: 22523539]
- Mueller SG, Laxer KD, Cashdollar N, Buckley S, Paul C, Weiner MW. Voxel-based Optimized Morphometry (VBM) of Gray and White Matter in Temporal Lobe Epilepsy (TLE) with and without Mesial Temporal Sclerosis. *Epilepsia*. 2006; 47:900–907. [PubMed: 16686655]
- Mueller SG, Laxer KD, Schuff N, Weiner MW. Voxel-based T2 Relaxation Rate Measurements in Temporal Lobe Epilepsy (TLE) with and without Mesial Temporal Sclerosis. *Epilepsia*. 2007; 48:220–228. [PubMed: 17295614]
- Mueller SG, Laxer KD, Barakos J, Cheong I, Garcia P, Weiner MW. Widespread neocortical abnormalities in temporal lobe epilepsy with and without mesial temporal lobe sclerosis. *Neuroimage*. 2009; 46:353–359. [PubMed: 19249372]
- Mueller SG, Laxer KD, Barakos J, Cheong I, Garcia P, Weiner MW. Subfield atrophy pattern in temporal lobe epilepsy with and without mesial sclerosis detected by high-resolution MRI at 4 Tesla: Preliminary results. *Epilepsia*. 2009; 50(6):1474–83. [PubMed: 19400880]

- Salmenpera TM, Symms MR, Rugg-Gunn FJ, Boulby PA, Free SL, Barker GJ, Yousry TA, Duncan JS. Evaluation of quantitative magnetic resonance imaging contrasts in MRI-negative refractory focal epilepsy. *Epilepsia*. 2007; 48:229–237. [PubMed: 17295615]
- Tao R, Fletcher PT, Gerber S, Whitaker RT. A variational image-based approach to the correction of susceptibility artifacts in the alignment of diffusion weighted and structural MRI. *Inf Process Med Imaging*. 2009; 21:664– 675. [PubMed: 19694302]
- Van Leemput K, Maes F, Vandermeulen D, Suetens P. A unifying framework for partial volume segmentation of brain MR images. *IEEE Trans Med Imaging*. 2003; 22:105–119. [PubMed: 12703764]
- Widjaja E, Mahmoodabadi SZ, Sneed OC III, Almelhaar A, Smith ML. Widespread cortical thinning in children with frontal lobe epilepsy. *Epilepsia*. 2011; 52:1885– 1891.

Highlights

Two-level multi-modality Bayesian network for classification of focal epilepsy.

1st level Bayesian network (GAMMA) identifies regions of GM and FA abnormalities.

2nd level Bayesian network uses information to provide group membership probabilities.

Accuracy to distinguish between 3 different epilepsy types was 64–87%

Figure 1.

Multi-modality second level Bayesian network. The network combines the information from the different imaging modalities (GM and FA) and the different comparisons at the first level. The dark orange nodes represent query nodes. Decision node 1 provides the probability of a subject being a Control, Focal Epilepsy or Focal Epilepsy with structural abnormalities not exceeding those in controls. Decision node 2 provides the probability of a subject with focal epilepsy being non-lesional temporal lobe epilepsy (TLE) or frontal lobe epilepsy (FLE) and Decision node 3 calculates the probability of a TLE subject having mesial-temporal sclerosis (TLE-MTS) or not (TLE-no). The yellow nodes are evidence nodes in which the information from GAMMA is entered and the bright orange nodes represent intermediary nodes. The edges are indicated by blue arrows.

Neuroimage. Author manuscript; available in PMC 2014 May 01.

Figure 3a

Figure 3b

Figure 3. GAMM maps of the different group comparisons. a. Belief maps of the various GM comparisons overlaid on the GM template. b. Belief maps of the FA map comparisons overlaid on the FA template. Please see text for a description of the distribution of the abnormalities.

Table 1a

Dice of Gray Matter Belief Maps

Overlap Map		No of Discriminating Voxels	TLE-MTS vs C	TLE-no vs C	FLE vs C	TLE-MTS vs TLE-no	FLE vs TLE-MTS	FLE vs TLE-no
Comparisons								
TLE-MTS vs C	82968			0.47	0.45	0.44	0.36	0.12
TLE-no vs C	92693	0.47			0.42	0.11	0.23	0.28
FLE vs C	62051	0.45		0.42		0.24	0.07	0.02
TLE-MTS vs TLE-no	85117	0.44		0.11	0.24		0.38	0.01
FLE vs TLE-MTS	70919	0.36		0.23	0.07	0.38		0.30
FLE vs TLE-no	39948	0.12		0.28	0.02	0.01	0.30	
<i>Mean</i>	<i>72282.67</i>	<i>0.37</i>		<i>0.30</i>	<i>0.24</i>	<i>0.24</i>	<i>0.27</i>	<i>0.15</i>
<i>SD</i>	<i>19222.67</i>	<i>0.15</i>		<i>0.15</i>	<i>0.20</i>	<i>0.18</i>	<i>0.12</i>	<i>0.14</i>
High Belief Overlap Maps		No of Discriminating Voxels	TLE-MTS vs C	TLE-no vs C	FLE vs C	TLE-MTS vs TLE-no	FLE vs TLE-MTS	FLE vs TLE-no
Comparisons								
TLE-MTS vs C	10818			0.11	0.11	0.39	0.37	0.00
TLE-no vs C	12700	0.11			0.08	0.04	0.06	0.07
FLE vs C	7267	0.11		0.08		0.06	0.01	0.00
TLE-MTS vs TLE-no	12545	0.39		0.04	0.06		0.35	0.00
FLE vs TLE-MTS	10725	0.37		0.06	0.01	0.35		0.01
FLE vs TLE-no	6244	0.00		0.07	0.00	0.00	0.01	
<i>Mean</i>	<i>10049.83</i>	<i>0.20</i>		<i>0.07</i>	<i>0.05</i>	<i>0.17</i>	<i>0.16</i>	<i>0.02</i>
<i>SD</i>	<i>2702.73</i>	<i>0.17</i>		<i>0.03</i>	<i>0.05</i>	<i>0.19</i>	<i>0.19</i>	<i>0.03</i>

TLE-MTS, temporal lobe epilepsy with mesial-temporal lobe sclerosis; TLE-no, temporal lobe epilepsy with normal MRI
 FLE, non-lesional frontal lobe epilepsy; C, control, please see text for the definitions of overlap map and high belief overlap map.

Table 1b

Dice of FA Belief Maps

Overlap Map							
Comparisons	No of Discriminating Voxels	TLE-MTS vs C	TLE-no vs C	FLE vs C	TLE-MTS vs TLE-no	FLE vs TLE-MTS	FLE vs TLE-no
TLE-MTS vs C	11224		0.23	0.12	0.30	0.02	0.15
TLE-no vs C	9336	0.23		0.25	0.02	0.06	0.03
FLE vs C	13034	0.12	0.25		0.22	0.30	0.46
TLE-MTS vs TLE-no	8573	0.30	0.02	0.22		0.03	0.35
FLE vs TLE-MTS	11366	0.02	0.06	0.30	0.03		0.44
FLE vs TLE-no	19877	0.15	0.03	0.46	0.35	0.44	
Mean	12235.00	0.16	0.12	0.27	0.19	0.17	0.29
SD	4065.23	0.11	0.11	0.13	0.15	0.19	0.19
High Belief Overlap Maps							
Comparisons	No of Discriminating Voxels	TLE-MTS vs C	TLE-no vs C	FLE vs C	TLE-MTS vs TLE-no	FLE vs TLE-MTS	FLE vs TLE-no
TLE-MTS vs C	1362		0.03	0.00	0.04	0.00	0.01
TLE-no vs C	1166	0.03		0.10	0.00	0.00	0.00
FLE vs C	1437	0.00	0.10		0.09	0.04	0.26
TLE-MTS vs TLE-no	1036	0.04	0.00	0.09		0.00	0.12
FLE vs TLE-MTS	1233	0.00	0.00	0.04	0.00		0.14
FLE vs TLE-no	2532	0.01	0.00	0.26	0.12	0.14	
Mean	1461.00	0.01	0.03	0.10	0.05	0.04	0.10
SD	543.52	0.02	0.04	0.10	0.05	0.06	0.11

TLE-MTS, temporal lobe epilepsy with mesial-temporal lobe sclerosis; TLE-no, temporal lobe epilepsy with normal MRI
 FLE, non-lesional frontal lobe epilepsy; C, control, please see text for the definitions of overlap map and high belief overlap map.

Table 2

Comparison of Classification Accuracy between Single and Multi-Modality Network

Type	Classification	GM alone	FA alone	GM and FA
TLE-MTS N = 19	% not Classified	15.8	10.5	0
	% Incorrectly Classified	5.3	36.8	15.8
	% Correctly Classified	78.8	52.6	84.2
TLE-no N = 22	% not Classified	31.8	27.3	18.2
	% Incorrectly Classified	22.7	54.5	22.7
	% Correctly Classified	45.5	18.2	59.1
FLE N = 14	% not Classified	57.1	35.7	21.4
	% Incorrectly Classified	7.1	14.3	28.57
	% Correctly Classified	35.7	42.8	50

Accuracies provided in Table are based on the total population, i.e. include subjects having abnormalities not exceeding those in controls (not classified) and are lower than those reported in the text that report the classification accuracy only for patients with diagnostic images, i.e., structural abnormalities exceeding those in controls.